DOCUMENT RESUME

ED 284 900 TM 870 451

AUTHOR Linn, Robert L.; Valencia, Shelia W.
TITLE Reading Assessments: Practice and Theoretical Perspectives. Research on Instructional Assessment: Instructionally Relevant Reading Assessment.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Aug 86
GRANT OERI-G-86-0003
NOTE 47p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; Decoding (Reading); *Diagnostic Tests; Elementary Secondary Education; *Informal Reading Inventories; Reading Comprehension; Reading Diagnosis; *Reading Tests; Semantics; *Standardized Tests; *Test Reviews

ABSTRACT

This paper evaluates widely used, standardized tests of reading within the framework of four interdependent reading components: (1) decoding speed and accuracy; (2) accuracy, fluency, and flexibility of determining semantic meaning of words; (3) sentence comprehension; and (4) passage dependent sentence comprehension. Among the types of standardized reading achievement tests are the group-administered survey tests, criterion-referenced testing systems, testing components of major basal reading series, individually-administered diagnostic reading tests, informal reading inventories, state assessment programs, and instruments of the National Assessment of Educational Progress. Types of score reports range from a single norm-referenced, global score to highly specific component scores based on only three or four test items. There is a relatively poor match between current theory and experimental research on the reading process and existing standardized tests of reading. Suggestions are provided for development of new measures of reading comprehension that are more consistent with modern theoretical and experimental research on the reading process. Eight pages of references are included. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view.

ED284900

Center for Student Testing, Evaluation and Standards

DELIVERABLE - AUGUST 1986

Research on Instructional Assessment: Instructionally Relevant Reading Assessment

Reading Assessments: Practice and Theoretical Perspectives

Robert L. Linn
Project Director

University of Illinois at Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Grant Number OERI-G-86-0003

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California

TM 870 451

The project presented, or reported herein, was performed pursuant to a grant from the Office of Educational Research and Improvement, Department of Education (OERI/ED). However, the opinions expressed herein do not necessarily reflect the position or policy of the OERI/ED and no official endorsement by the OERI/ED should be inferred.

Reading Assessment: Practice and Theoretical Perspectives

Robert L. Linn and Sheila W. Valencia
Center for Student Testing, Evaluation, and Testing
University of Illinois at Urbana-Champaign

No area of student learning is more fundamental to student progress throughout school than reading. Over the last decade there have been great advances in the scientific understanding of the reading process (National Academy of Education, 1985). These advances are beginning to have an impact on the design of instructional materials and approaches to teaching reading (Pearson, 1986). To date, however, they have had little impact on testing and there is a relatively poor match between theory and experimental research on the reading process and standardized reading tests.

This apparent lack of match between reading research and approaches used to assess reading comprehension provides the focus for this review. As was noted by Curtis and Glaser (1983), recent research on the cognitive processes involved in reading suggests that there are four major, interdependent components of reading comprehension. These are (1) decoding speed and accuracy, (2) accuracy, fluency, and flexibility of determining the semantic meaning of words, (3) passage dependent sentence comprehension, and (4) passage comprehension which involves the development of a "coherent cognitive model of text meaning" (Johnston, 1984, p. 236).

The four interdependent reading components identified by Curtis and Glaser provide the framework for this review. Existing, widely used, standardized tests are evaluated within this framework and

suggestions for the development of new measures of reading comprehension that are more consistent with modern theoretical and experimental research on the reading process are provided.

Variety and Uses of Reading Tests

Clearly, there is no shortage of standardized reading tests available for use in the schools. Almost all school districts administer standardized reading achievement tests each year. Included among the standardized tests in use are (1) group-administered survey tests - either stand-alone tests such as the Gates-MacGinitie Reading Tests (MacGinitie, et al., 1978) or parts of batteries such as the Comprehensive Tests of Basic Skills (CTB/McGraw-Hill, 1982); (2) criterion-referenced testing systems, e.g., the PRI Reading Systems (CTB/McGraw-Hill, 1980); (3) the testing components of major basal reading series, e.g., Macmillan's Testing and Management Resource Books (Weinstein, 1983); (4) individually-administered diagnostic reading tests that are intended to provide more detailed informaion about a student's strengths and weaknesses, e.g., the Diagnostic Reading Scales (Spache, 1981); (5) informal reading inventories, e.g., the Basic Reading Inventory (Johns, 1981); (6) state assessment programs; and (7) the National Assessment of Educational Progress.

The types of score reports are as variable as the types of tests, ranging from a single norm-referenced, global score to highly specific component scores that may be based on only three or four test items. The oral reading sections of some tests, for example, provide information on a variety of student errors such as additions, substitutions, omissions, mispronounciations, and reversals as well as indications of word recognition and comprehension of text. Separate scores for initial, middle, and final constants; long and short

vowels; consonant blends and diagraphs may be among the detailed scores that can be obtained.

Detailed information on student performance on specific skills is also commonly provided by "mastery tests" that have become a standard part of the curriculum materials sold by the publishers of basal readers. The level of detail on these curriculum-embedded tests is usually very fine. For example, short tests for specific objectives such as "recognize the phoneme-grapheme correspondences of dipthongs" or "divide for syllabication a two-syllable word with medial consonant letters" (Smith & Arnold, 1983) can be found in assessment systems accompanying basal readers.

Theoretical Foundations

Despite the wide array and widespread use of tests of reading, considerable dissatisfaction with the available tests has been expressed by the reading reasearch community (Valencia & Pearson, 1986). Some of the dissatisfaction is illustrated by Pearson and Herman's (1985) review of the Gates-McKillop-Horowitz Reading Diagnostic Test, Second Edition (1981).

The Gates-McKillop-Horowitz test is intended to provide diagnostic information on a number of word recognition skills for students in grades 1 thorugh 6. Twenty three scores, for characteristics such as omissions, mispronunciations (e.g., wrong beginning, wrong middle, accent errors), and recognizing and blending common word parts, are provided. Nonsense words, which are purportedly made up of two or more frequently used syllables, are used on the syllabication subtest.

As noted by Pearson and Herman (1985), the test follows a

traditional view of reading as composed of a a series of subskills and presumes that the separate assessment of each subskill will provide diagnostically useful information. It is precisely this view of reading, however, that many reading researchers reject. Pearson and Herman (1985) express their concern in this regard as follows:

> "To assess sheer knowledge of these 'pieces' outside the context in which the student is expected to apply that knowledge, i.e., when the student is really reading real text, gives a biased estimate of ability to use that knowledge. Not only does assessment devoid of context prevent the student from using the rich range of resources available in most real text, it also prevents the examiner from getting a picture of how the student is or is not able to marshall resources, skills, and strategies to solve the problem of what the text means" (p. 602).

Similar concerns have been expressed by other reading researchers about the emphasis on a host of discrete skills in many criterion-referenced reading test systems. In her review of the PRI Reading Systems (CTB/McGraw-Hill, 1980), for example, Hiebert (1985) concluded that the "assessment program suffers from the major deficiencies that have typically characterized criterion-referenced programs ... The most fundamental issues involve the assumptions that mastery of these objectives constitutes reading and that these objectives are acquired in this particular order" (p. 1201).

The emphasis on discrete skills in inconsistent with the emerging conception of reading as an integrative process (National Academy of Education, 1985; Stanovich, 1980). So too are the emphases found in most individually-administered diagnostic reading tests. Indeed, there is little theoretical rationale to support these tests, and even

less in the way of evidence that the putatively diagnostic information that is provided leads to valid instuctional decisions.

According to the National Academy of Education's Commission on Reading (1985), "[r]eading is the process of constructing meaning from written texts. It is a complex skill requiring the coordination of a number of interrelated sources of information" (p. 7). Skills such as the ability to decode words and to assess the semantic meaning of a word are certainly necessary for the skilled reader, but reading comprehension requires more than the application of a series of discrete skills. Reading is an active process that requires the integration of information provided by the text with the reader's prior knowledge (e.g., Bransford & Johnson, 1973; Johnston, 1983; Pearson & Spiro, 1980). Purposive reading requires self-monitoring and self-regulatory skills that have come to be known as "metacognitive abilites (e.g., Brown, 1978; Flavell, 1976).

Higher level integrative and metacognitive abilities receive relatively little attention on current tests which emphasize discrete skills. Instead, tests often emphasize literal comprhension rather than inference and rarely go beyond asking a student to find the main idea of a brief passage (Valencia & Pearson, 1986). As noted by Curtis and Glaser (1983), both reading instruction and achievement testing, although adequately addressing lower level reading requirements, may be doing so at the expense of higher levels" (p. 133). They go on to argue that "an alternative view of reading instruction and testing is needed in which (a) the skills involved in reading are assumed to be more interrelated than in a stage model, and (b) revisions are made in a way that teaching and testing take into

account interactions among components of reading" (Curtis & Glaser, 1983, p. 133).

Recent research in cognitive psychology (see, for example, Curtis & Glaser, 1983; Johnston, 1983; Stanovich, 1980) provides a theoretical foundation for the evaluation of current reading tests and suggestions of new approaches to the measurement or reading comprehension. Curtis and Glaser (1983) clearly illustrated how theory and research on cognitive processes in reading can be used to specify desirable characteristics of tests of reading comprehension. As was noted above, their analysis led to the identification of four major, interrelated components of reading comprehension: (1) decoding speed and accuracy, (2) accuracy, fluency, and flexibility of semantic meaning of words, (3) passage dependent sentence comprehension, and (4) passage comprehension.

Although each of these components is important for skilled reading and may be measured separately, the key aspect of this framework is the emphasis on the integration of the components. Perfetti and Lesgold (1979) illustrated the essential interdependence of the components of reading by analogy to the components of high fidelity systems. "If any component of the system is defective, sound quality suffers. The components can be independently tested and, more importantly, independently improved. However, improvement of one component may not immediately improve sound quality (but it may increase the potential of the system to benefit from later improvements in other components)" (Perfetti & Lesgold, 1979, p. 58). In an analogous fashion, it is argued that the components of reading comprehension must be integrated to produce meaning from text.

Perfetti and Lesgold's hi-fi analogy forcefully illustrates the

key role that reading theorists attach to the integrative aspects of reading. As is true of most analogies, however, it can be carried too far. There is evidence, for example, that the components of reading, unlike those of a hi-fi system, are not only interactive but are compensatory. Stanovich (1980) has reviewed a large body of theoretical and experimental work that suggests the need for what he has referred to as "an integrative-compensatory model of individual differences in reading ability" (p. 63).

The need for a compensatory model is best illustrated by the reliance on context to identify words. Skilled readers apparently rely little on context to identify words (McConkie & Zola, 1981) because their word recognition and decoding skills are so efficient that there is seldom need to depend on the context to identify individual words. Top down models of reading that hold that readers are continually hypothesis testing as they read are implausible because they require that readers are able form and test hypotheses "based on complex syntactic and semantic analyses ... in less than the few hundred milliseconds that is required for a fluent reader to recognize most words" (Stanovich, 1980, p. 34). If anything, the evidence suggests that less-skilled readers rely more on context for word identification that do more-skilled readers (see Stanovich, 1980, for a review of several studies supporting this conclusion).

The use of context by readers with poor decoding skills illustrates the compensatory nature of reading. It is not that good readers lack the ability to use context. Indeed, there is evidence that good readers tend to be able to predict words that are missing from a passage better than poor readers. However, it does not

necessarily follow that because good readers have superior predictive abilities that they rely on these ablities to recognize words (Stanovich, 1980). Rapid word identifcation by skilled readers occurs automatically and frees cognitive resources for higher level processing and interpretation of the text (Curtis & Glaser, 1983). It seems likely that it is simply more efficient for readers "with well developed decoding skills to directly identify words in a test-driven manner than to 'predict' words based on context" (Juel, 1980, p. 49).

Readers with less ability to recognize individual words automatically may compensate for this deficit by placing greater reliance on context and stored knowledge. Such compensation is only partially effective, however, because it comes at a cost. It requires the use of cognitive resources that are needed for the higher-level processing that is essential for the construction of meaning from the text. Thus, as in Perfetti and Lesgold's analogy to a hi-fi system, a weak component can limit the overall quality of the system because the other components can't work to capacity, much less, fully compensate for the weak one.

The four interrelated components of reading identified by Curtis and Glaser (1983) provided a framework for reviewing currently used tests in the following sections of this report. Given the large number of published tests and the substantial degree of similarity among many of the tests that are designed to serve a given purpose, an exaustive review of all published tests was not attempted. Such a review would be highly redundant. Hence, major examples that illustrate the range of approaches in each area were reviewed. For each component, an attempt was made to analyze the tests in terms of recent research on reading and to use that research as a basis for

suggesting future directions for the testing of these components.

Decoding Speed and Accuracy

As has already been suggested, word decoding is a key component of the reading process - one of several interdependent processes that contributes to skilled reading. The work of Samuels and LaBerge (1974) on automatic processing of words has suggested that automatic decoding of words is a key to the reduction of attentional demands of the reading task. When these demands of word tasks are reduced, more attention is freed to be allocated to other, and perhaps higher level, processes such as constructing a coherent model of the text.

As summarized by Stanovich (1980), speed of word recognition is important because rapid processing of information in short term memory enables the reader to utilize higher level integrative comprehension processes that operate on the information stored there (Lesgold & Perfetti, 1977; Perfetti & Lesgold, 1977). Numerous studies have suggested that rapid, context-free word identification is a major factor distinguishing skilled readers from less skilled readers and that speed of word recognition continues to increase after the leveling off of automaticity at the second or third grade levels (Curtis, 1980; Perfetti & Hogaboam, 1975; West & Stanovich, 1979; Guttentag & Haith, 1978; Hogaboam & Perfetti, 1978).

The available evidence strongly supports the conclusion that the ability to identify words accurately and quickly is critical to becoming a skilled reader. Although some readers are able to recognize words adequately, they respond so slowly that meaning is often lost. At the other extreme, there are readers who will "read" words very quickly but may fail to get any meaning from them because

the words are incorrectly identified. In fact, stages of sight word acquisition (Adams & Higgins, 1985; Samuels & LaBerge, 1974; McCormick & Samuels, 1979) suggest that students first must develop accuracy with word identification before they become automatic or increase speed of identification. While some studies have found significance when examining reading ability using only accuracy of word identification (Adams & Huggins, 1985; Juel & Roper-Schneider, 1985), others have focused, and found significance when looking at speed of word processing alone (Biemiller, 1977-78; Hogaboam & Perfetti, 1978; Perfetti, Finger & Hogaboam, 1975).

Some researchers, however, have investigated good and poor readers in terms of speed and accuracy of word identification. For example, Shankweiler & Liberman (1972) indicated that word list accuracy for students in grades 2 to 4 was highly correlated with reading performance on paragraphs. However, after examining the correlations between latency, accuracy and paragraph reading for one group of third graders, they concluded that slow rate of individual word identification may contribute as much as inaccuracy to poor performance on paragraph reading.

There are numerous individually-administered tests that assess the recognition of individual words. The word recognition section of the Wide Range Achievement Test (Jastik & Jastik, 1978) is illustrative of the general approach that is often used. Lists of letters and words are presented to children and the accuracy of pronunciation is recorded. On Level I of the WRAT, which is intended for use with children aged 5 through 11, the words are arranged approximately in order of increasing difficulty and testing is dicontinued after 10 consecutive errors.

The word recognition section of the WRAT is easy to administer and yields scores with relatively high reliability. It does not provide a measure of speed of word recognition, however. Furthermore, the accuracy scores are interpretable only in norm-referenced terms. This is so because there is no sound rationale or systematic basis for the selection of words for the test.

Several informal reading inventories include word recognition sections. The Basic Reading Inventory, Second Edition (Johns, 1981), for example, includes a "word recognition in isolation" subtest. The word list is said to be graded and may be administered as a timed or untimed test. Although it has been suggested that the comparison of timed and untimed performance may by useful (Plessas, 1985), speed of vocalization is not directly assessed. The basis for the selection of words is unclear and, as is typical of several informal reading inventories, no statistical support of the reliability or validity of the test is provided. Nor is the evidence provided that would justify the use of the results on the word recognition section to determine initial placement in the test's reading passages.

The Ekwall Reading Inventory (Ekwall, 1979), a similar informal reading inventory, uses a graded word list for determining initial placement into reading passages. As was true of the Basic Reading Inventory, there is a lack of technical information or justification of the choice of words for the word recognition portion of the inventory. The intended use of the word recognition in isolation sections of tests such as the Ekwall and Basic reading inventories is relatively limited, however.

Some published tests focus exclusively on word recognition. Two

such tests are the Instant Words Criterion Test (Fry, 1980) and the Pope Inventory of Basic Reading Skills (Pope, 1974). The Instant Word Criterion Test consists of 300 common words, which the author claims are the most common words in English. As noted by Elkins (1985, p. 692), "[n]o references are given to support this claim, but an inspection suggests that these words are indeed basic sight vocabulary." It is presumed that these words should be recognized instantly. Children are required to read the list of words and the examiner records words that are "missed". Speed of recognition is not assessed and the criteria for recording a failure are not specified. "Neither is there any indication given of how much time should be allowed to the pupil to answer each item, which is surely important in view of the word 'instant' in the title of the test and the implied assumption" (Carroll, 1985). It is simply assumed that words that are missed should be taught since recognition of these words is needed for reading. Thus, only item scores are suggested. There is no supporting evidence of reliability (either agreement among examiners or test-retest) or of validity.

The Pope Inventory of Basic Reading Skills is an individually administered test that is intended to assess word attack skills. Section II of the inventory is used to assess a child's basic sight vocabulary by presenting the child with a deck of cards with one common word printed on each card. The child is instructed to first sort the cards into two piles: known words and unknown words. The child is then asked to read aloud the "known" words and the number read correctly is used as an estimate of the child's sight vocabulary. No norms, reliability, or validity data are provided.

The Biemiller Test of Reading Processes (Biemiller, 1981) is one

of the few published tests that attempts to measure speed of word recognition directly. Speed in naming letters, words in isolation, and words in context is measured. The words in isolation part of the test is divided into two lists of 50 words each. The first list is drawn from primer-level vocabulary and the second from middle-elementary school level vocabulary. Percentile scores (90th, 75th, 50th, 25th, or 10th) are reported. The percentiles, however, are based on a sample of 340 children from 17 Canadian schools, and are not representative of any clearly defined population.

The precise basis for sampling words to be included on the lists is unclear. Although the directions for administration are quite specific, it is likely that considerable practice would be needed to obtain accurate measures of speed. The following comments by Fry (1985) are worthy of note in this regard.

> "The problems of accurately recording reading time ... indicate the examiner would need some training and experience. It is important to keep the child reading. For example, the author states, `If the child pauses over a word or rereads a line, stop the stopwatch until the child is reading words she/he has not read before.' In addition, `If there are more than three delay-type errors (as opposed to misreadings without stopping), the data will be invalid, and testing should be stopped.' This is a lot for a novice to remember when seconds count" Fry, 1985, p. 194).

Although not really a test, the IOX Basic Skills Word List (IOX Assessemnt Associates, 1980) deserves mention as the final instrument in this section. As the name suggests, the IOX Basic Skills Word List

is a list of words. A total of 7,318 words are included in the list which is orgaized by grade level (from 1 to 12). The words were selected using information from several sources using the following criteria: "(1) the frequency with wiich words occur in reading textbooks, (2) the frequency of words in generally read materials, and (3) children's demonstrated familiarity with particular words" (IOX Assessment Associates, 1980).

The initial list and grade level assignments were taken from Taylor, Frackenpohl, White, Nieroroda, Browning, and Birsner's (1979) core vocabulary list. Modifications of grade level assignments and additions to the list were based on familiarity of words to children at different grades as indicated by Dale and O'Rourke (1976) and Sakiey and Fry's (1979) 3000 most frequent words from the Carroll, Davies, and Richman's (1971) Word Frequency Book.

Though not a test in the usual sense, it is suggested that the IOX Basic Skills Word List might be used to construct informal tests of word recognition for use in "diagnosis". The list is also intended as a resource for judging the readability level of texts or passages to be used on tests and for identifying words for "direct teaching". The rationale for the suggested diagnostic use in unclear, however. More importantly, the notion that words should be identified in this manner for direct teaching is not justified. The heavy emphasis on individual words at the expense of time devoted to reading complete stories or text is at best debatable (see, for example, Anderson & Freebody, 1983).

This brief review of word recognition measures leads to two generalizations: (1) speed of word recognition is only rarely measured directly and (2) the selection of words for tests is often more of an

art than a science. There is seldom a clear justification for the inclusion of particular words on a test. But the selection of words for tests of speed and accuracy of word identification is certainly critical to meaningful interpretation of results, in other than a norm-referenced sense, and to relevant classroom instruction. Our review of the research suggests a number of factors that are potentially relevant in the classification of words, including word frequency, inclusion and emphasis in the curriculum, the approach of the instructional program, orthographic complexity, word type, and word length. Based on our review, however, we have concluded that the three factors that are most critical for creating specifications for a test of speed and accuracy of word identification are (1) word frequency, (2) exposure to words in the curriculum, and (3) the instuctional program.

Skilled readers are faster than less-silled readers at identifying high frequency words (Biemiller, 1977-78; Curtis, 1980; Perfetti, Finger & Hogaboam, 1978; Perfetti & Hogaboam, 1978; West & Staovich, 1979). At the same time, however, Perfetti & Hoagaboam (1975) have demonstrated that differences between good readers and poor readers are even greater for low frequency and pseudowords. This suggests that speed is not just familiarity with particular words. In fact Juel & Roper-Schneider (1985) found that first grade students with a more intensive phonics instructional program were more successful at correctly pronouncing words not seen before than students in a more visual word recognition program.

West and Stanovich (1979) have cautioned that one of the reasons researchers may have been unable to demonstrate developmental trends

in automaticity is that many of the words used on measures have been too easy. That is, there is a sort of ceiling effect. This same caution is applicable in the construction of lists of words for tests.

Related to the issue of frequency, is exposure to words or the curricular validity of those words. Juel & Roper-Schneider (1985), for example, found that the types of words in first grade texts exert a more powerful influence on children's word recognition strategies than the method of instruction. Additionally, they found that repeated exposures to words, the number of repetitions in the basal, was a significant factor in accuracy of basal word identification. Both these findings suggest that increased reading may provide added exposure to words that may in turn increase word identification speed and accuracy. These findings also suggest that the match between the words on a test and the textbooks and instructional program materials may be important determinates of the instructional validity of a test.

Semantic Meaning

Although the research reviewed above makes a strong case for the importance of being able to recognize words accurately and quickly, this ability represents only one of the interrelated components needed by a skilled reader. The breadth and depth of an indivdual's vocabulary knowledge is also vital and provides a good predictor of reading comprehension ability (Thorndike, 1973). It is almost self evident that a reader who does not know the meaning of a large number of words in a passage will have great difficulty comprehending it. There is considerable evidence to support this common sense conclusion (see, for example, Anderson & Freebody, 1981). However, as Curtis and Glaser (1983) have cautioned, simply learning word meanings "does not always lead to comprehension improvement (Jenkins, Pany, & Schreck,

1978; Tuinman & Brady, 1974). For vocabulary instruction to have an effect on reading comprehension, it needs, at a minimum, to go beyond dictionary definitions to deep understandings and related ideas (Draper & Moeller, 1971).

Anderson and Freebody (1983) have made a convincing case that estimates of vocabulary size are potentially important both for purposes of policy decisions and for purposes of assessing individual differences. As they demonstrate, however, there are order-of-magnitude differences in the existing estimates of vocabulary size. Anderson and Freebody cite two reasons for the widely divergent estimates. First, there are differences in the definitions of the domain of potential words, i.e., what counts as a separate word and how inclusive the word list is for which estimates are sought. Second, the estimates depend on the methods used to measure knowledge of selected words.

Reading tests commonly report separate scores for vocabulary knowledge, but they do not purport to give an estimate of a test taker's vocabulary size. Rather, intrepretations are generally based on normative comparisons. Hence, before considering possibilities for obtaining estimates of an individual's vocabulary size we will review some of the major measurement approaches used on standardized tests of vocabulary knowledge.

A variety of formats are used on tests of vocabulary knowledge. By far the most common procedure, however, is to use multiple choice questions. Either isolated words or words in context are commonly presented and the test taker is instructed to select one of four or five options that has the same or most similar meaning. When words in

context are presented, the context may consist only of a two or three word phrase or one or more complete sentences depending on the test. At higher grades antonyms, rather than synonyms are also frequently used.

Given the high degree of similarity among tests of vocabulary knowledge used in the elementary grades, only a few tests will be described here. The Gates-MacGinitie Reading Tests (MacGinitie, 1978) is a well known and widely used reading test series. It consists of seven levels that together span grades 1 through 12. As is typical of such series or of comprehensive achievement test batteries, vocabulary as well as comprehension is assessed at each level. In the lower grades, vocabulary is assessed by having test takers select one of four words that describe a picture. Starting at grade three, test takers are required to select a word or phrase that has the most similar meaning to the word presented in the item stem. Using a variety of sources, "[w]ords were selected to characterize those likely to be found in reading materials in the grade range covered by each test level" (Rupley, 1985, p. 595). The test has norms both for in-level and out-of-level testing and high internal-constistency reliabilities (.90 to .95).

The Word Meaning test of the Nelson Reading Skills Tests, Forms 3 and 4 (Hanna, Schell, & Schreiner, 1977) uses two types of items, both of which require the test taker to select the nearest synonym from four options. The stem consists of either isolated words or words in phrases. The vocabulary for the tests was selected from humanities, social studies and science texts. Norm-referenced scores with good split-half reliabilities (.89 to .93) are reported.

The vocabulary section of Level E of the Comprehensive Tests of

Basic Skills, Forms U and V, (CTB/McGraw-Hill, 1982), which is typically used in the third grade, uses three item formats and illustrates some of the range of items types found on widely used tests of vocabulary knowledge. The first item type presents a two or three word phase with one underlined word (e.g., "very large") as the stem and the test taker is told to select the "word below that means the same or about the same as the underlined word". For the above sample item, for example, the the options are "big", "old", "brave", and "strong". The stem for items in the second section presents two underlined phrases (e.g., "a baseball player and something used for pouring") and the child is instructed to pick the word that fits both ("glass", "bottle", "catcher", or "pitcher"). The third item type has one or more sentences in the stem in which one word is underlined (e.g., "Billy sat crying on the steps. His dog was lost. He was so miserable."). The child is instructed to select one of the four optional words that means the same or nearly the same as the underlined word ("cold", "sleepy", "excited", "unhappy").

These items illustrate that tests of vocabulary may tap a variety of kinds of knowledge about a word. As noted by Curtis and Glaser (1983, p. 137), "vocabulary items differ in the extent to which they assess individual's abilities to: (a) recognize a correct meaning of a word; (b) determine which of several correct meanings is appropriate in a particular context; and, in some cases, (c) figure out an unknown word's meaning from context." A given child might use all three of these strategies in answering the three sample items quoted above from the CTBS. Clearly, it could not be safely assumed that a correct answer for the word "miserable" implies that the child would also have

answered in the same way if the word were presented in isolation. The word miserable is a low frequency word in third grade materials, ocurring only twice in the over 800,000 words from third grade texts analyzed by Carroll, Davies, and Richman (1971). Yet with the supporting context it was judged to be easy enough to use as a sample item.

Anderson and Freebody (1983) report a number of fairly wide discrepencies between the percentage of children who claim to know a word and the number who give the correct answer to a multiple choice item involving that word. The discrepency can go in either direction, depending on the nature of the multiple choice item. For example, only 19% of their sample of fifth grade students said that they knew the word "judicious", yet 51% of the sample gave the correct answer to the following multiple-choice item:

"A judicious decision is made -

1. quickly 3. foolishly

2. wisely 4. cleverly"

On the other hand, 96% of the sample said they knew the word "manage", while only 28% gave the correct answer to the following multiple-choice item:

"If you manage on your allowance, you -

1. spend it 3. get along

2. save it 4. waste it"

Anderson and Freebody found a closer correspondence between interviews of students about their knowledge of word meanings and a simple yes or no test of whether or not a student knew a word than with the results of multiple choice items. They concluded "that when the word tested in a standardized multiple choice item is difficult

something about the item will tend to give away the correct answer, whereas when an easy word is tested the item will tend to lead the student away from the correct answer" (Anderson & Freebody, 1983, p. 238). This tendency is likely to be the consequence of reliance on item analysis statistics in the selection of test items. On norm-referenced tests, preference is traditionally given to items of middle difficult. Extremely easy and extremely difficult items are excluded. Thus, for an easy word to be included fine distinctions may be required for the item to pass the item analysis screen, whereas the converse may be true for difficult words.

Indirect support for Anderson and Freebody's conclusion is provided by a comparison of item difficulties (as indicated by the location parameter from the three parameter logistic model) and the frequency of occurrence of the target words. In general one would expect a substantial relationship between the frequency of occurrence of words in books and knowledge of word meanings. However, for the 30 target words in the vocabulary section of Level E of the CTBS Form U, the correlation between the Standard Frequency Index found in Carroll, Davies, and Richman (1971) and item difficulty (location parameter) is only -.09. Thus, there is only a slight tendency, if any, for the low frequency target words to be more difficult than the high frequency words. This poor relationship is likely to be attributable to the requirement of making finer distinctions for high frequency words (e.g., "famous" and "whole") than low frequency words (e.g., "chuckle" and "thaws"). More clues are also apt to be given by the context accompanying low frequency words.

For norm-referenced interpretations or for purposes of

prediction, the ambiguity of the information provided by a multiple choice item about a student's knowledge of the meaning(s) of a word is unimportant. However, if the goal is to estimate vocabulary size or to have an indication of whether or not students know the meanings of words found in instructional materials, this ambiguity is undesirable.

An alternative approach to measuring word knowledge has recently been investigated by Anderson and Freebody (1983). They used a simple yes/no method where a student is required only to indicate the words he or she knows by means of a check or some other device (e.g., pressing a button). This approach eliminates extraneous factors such as those due to context or the nature of the multiple choice options that are selected. The obvious drawback is that students may differ in their propensity to respond "yes" to words that they don't know or to respond "no" to words about which they are unsure.

To guard against these possibilities, Anderson and Freebody intersperse pseudowords with the real words on the test. The proportion of words known is then estimated by the difference in the proportion of yes's to words (hits) and the proportion of yes's to pseudowords (false alarms) all divided by one minus the proportion of false alarms. A large number of words and pseudowords can be presented in a fairly short period of time using this techinique. While not without difficulties (e.g., the problem of specifying the rules for sampling words and for constructing pseudowords), the method has considerable appeal for some purposes.

Anderson and Freebody's yes/no method may prove useful for obtaining criterion-referenced estimates of some, but not all, aspects of a student's word knowledge. Curtis and Glaser (1983), for example, have suggested that there are three aspects of semantic word meaning

that may be important to distinguish for purposes of measurement. These are accuracy, flexibility, and fluency. Accuracy "reflects whether or not there is any appropriate semantic knowledge in memory associated with a word" (Curtis & Glaser, 1983, p. 136). Flexibility refers to the depth of an individual's knowledge of word meaning and fluency refers to the speed of access to the meaning of a word. As in the case of slow decoding, slow access to the meaning of a word uses cognitive resources that are needed for processing sentences and comprehending longer segments of text. Selecting the most appropriate of several meanings of a word is essential to comprehension of the meaning of a sentence as a whole and the adequacy of the overall comprehension of the meaning of the text can depend on the richness or depth of knowledge about particular words.

Current published tests of word meaning do not distinguish among the aspects identified by Curtis and Glaser. The lack of such distinctions is of no real concern if the goal of measurement is to rank order students or to predict future performance, but the distinctions could prove important for purposes of planning instruction. Anderson and Freebody's yes/no procedure may provide a useful alternative for assessing accuracy and possibly fluency. Other, more adaptive techniques, are likely to be needed for obtaining measures of flexibility, however.

Sentence Comprehension

The third of the four components of reading suggested by Curtis and Glaser (1983) was called sentence processing. Sentence processing refers to the "integration of each incoming sentence into the memory structure that exists for what has already been read" (Curtis &

Glaser, 1983, p. 138). Such integration may involve a simple matching of words or concepts in one sentence with those in the immediately preceding sentence. For example, a pronoun may be matched with a noun in the previous sentence. New information contained by the second sentence may provide an elaboration of a concept in the first sentence. Comprehension in this case requires an integration of the new information with the concept in short-term memory.

Integration involving simple matching of words or the linking of new information to elaborate a previous concept in short-term memory can be illustrated by the first two sentences of the first reading passage of the Elementary Level (grades 3.5-4.9) of the Meteropolitan Achievement Tests, Fifth Edition, Form JS (Prescott, Balow, Hogan, & Farr, 1978).

"Our block used to have many oak trees along the street.
It was quite shady and cool in the summer."

The second sentence is linked to the first by the pronoun "it" and the concept of the oak trees is elaborated by "shady and cool".

Of course, the information in a new sentence does not always involve such direct matches to information contained in short-term memory. When it doesn't, the reader must rely on other integrative processes. This may involve searching long-term memory for information encountered earlier in the passage or the prior knowledge the reader brings to the task. Curtis and Glaser (1983) used the following four sentences to illustrate the need to reactivate information stored in long-term memory.

"A thick cloud of smoke hung over the forest. Glancing to the side, Mary could see a bee flying around the back seat. She pulled off the road and rolled down her window. The forest was

on fire" (pp. 138-139).

Because the second and third sentences deal with a topic different than the first, the "cloud of smoke" is no longer likely to be in short term memory. Reactivation of this information is signaled by the reintroduction of the word "forest" (Curtis & Glaser, 1983). Such integration of information across sentences is critical to a good understanding of the text.

Lacking prerequisite prior knowledge or having an inadequate memory of the previous information supplied by the passage reduces comprehension. "Both the efficiency of decoding skills and the sufficiency of semantic word knowledge can affect the manipulability and availability of passage information. Individuals who have such reading difficulties tend to be less able to hold specific words from a passage in short-term memory (Perfetti & Goldman, 1976)" (Curtis & Glaser, 1983, p. 139).

Sentence comprehension is frequently assessed in standardized reading tests by presenting questions that paraphase parts of the text. A large proportion of questions on some tests can be answered by direct comparison of the question to a sentence in the text without necessarily being able to understand the sentence (Anderson, 1972; Curtis & Glaser, 1983). It has also been demonstrated (e.g., Tuinman, 1974) that a substantial fraction of questions on some standardized reading tests can be answered correctly without reading the text based simply on prior knowledge. To assess integration, it is important that questions (1) be passage dependent, i.e., not be answerable on the basis of prior knowledge without reading the target sentences and (2) require understanding rather than being answerable on the basis of

word matching.

The better standardized tests have attempted to address these concerns. Questions that require integration of information from different sentences and that have a low likelihood of being answered correctly without reading the passage have been included with increasing frequency in revisions of the widely used standardized reading tests. The passage form the Meteropolitan cited above can be used to illustrate this type of question. The end of that passages contains the following four sentences:

> "But last year, some people came in huge truck and chopped down all the oak trees. They wanted to make the street wider. I hope the squirrels and the birds will return to our block. The other children want them back too, so today we planted six small trees."

One of the questions following this passage is:

> "After the trees were cut down, the children missed the -
>
> A. trucks C. men
>
> B. animals D. noise"

Without the passage, all of the options are plausible, i.e., it is reasonable to expect that the question is passage dependent. The question also requires an integration of the concept of "animals" with "the squirrels and the birds" and an inference that "I hope" and the "other children want" because the squirrels and birds were "missed". On the other hand, the child who depends heavily on prior knowledge and personal preference may reasonably infer that the children missed the trucks more than the animals. As noted by Valencia and Pearson (1986), individual differences in prior knowledge can lead to differences in the inferences that are made from a given text, and

more than one inference is often defensible.

The Degrees of Reading Power (Touchstone Applied Science Associates, 1983) illustrates another approach to testing the integration of information across sentences. The DRP has features similar to a cloze test. Words are deleted from a passage and the test taker has to select one of five options that should be used for the deleted word. It differs from a typical cloze test, however, in several important respects. First, fewer words are deleted (only seven per passage which typically run approximately 325 words in length) than on the usual cloze test. Second, the deleted word and words in the sentence where it belongs are all familiar, relatively high frequency words. That is, in comparison to the surrounding sentences, the target word and sentence is easy. Thus, the dependence of the DRP scores on the particular vocabulary of the response options is minimized. Third, if only the target sentence is considered, all of the options are reasonable. This is an essential part of the DRP approach. It is intended to ensure "that processing surrounding prose is both necessary and sufficient to choosing the right answer" (Koslin, Koslin, & Zeno, 1979, p. 316).

The following sample item from Form PX-1 of the DRP illustrates this key third feature.

"It was sunny and hot for days. S-1 a) price b) road
Then the __s-1__ changed. c) job d) weather
It turned cloudy and cool. e) size"

Although shorter than the actual test passages, it can be seen that there is no good basis for choosing among the four optional words when the target sentence is considered in isolation. When integrated with

the information in the preceeding and subsequent sentences, however, only "weather" leads to a coherent set.

The examples from the Metropoitan and the DRP show that existing test items can and do tap some of the integrative aspects of sentence processing that were highlighted by Curtis and Glaser (1983). Other aspects, however, seem to be missed. For example, since the passage is available for reference when the questions are considered, it is not necessary to rely as heavily on the retrival of concepts from long-term memory as may be necessary in skilled reading. Fewer demands may also be made on the manipulation of concepts in short-term memory. Certainly, a test on which the text was unavailable would change the task demands. Whether such a procedure would provide additional useful information about individual differences in reading ability is an open question, however.

Integrative processing serves as a basis for detecting and recovering from misinterpretations of text. Markman (1977, 1979) has demonstrated that children who have difficulty in understanding text, even text containing only simple words that the children can readily decode and that are familiar in meaning, are frequently unable to detect inconsistencies. This is illustrated by one of his examples: "Fish must have light in order to see. There is absolutely no light at the bottom of the ocean. ... Some fish who live there know their food by its color" (Markman, 1979, p. 646). Failure to detect the inconsitencey is symptomatic of the child's lack of integration of the information contained in the third sentence with the information provided earlier by the first sentence. The use of inconsistencies such as the one illustrated above represents a promising approach to the construction of reading tests that assess a child's ability to

integrate information across different parts of a text.

Passage Comprehension

According to Anderson (1977, p. 419) "[t]o comprehend a message is to discover a formulation which coherently explains its contents." Comprehension of the message in a text requires more than the ability to decode and attach meanings to words. It requires more than the ability to recognize paraphrases of the sentences. It involves the "forming of a coherent cognitive model of the text meaning" (Johnston, 1984, p. 236). Development of an appropriate cognitive model depends on the schema that the reader possesses as well as the properties of the text (e.g., Anderson, Reynolds, Schallert & Goetz, 1977; Bransford & Johnson, 1973; Rumelhart & Ortony, 1977).

Schema theory attempts to explain how knowledge is stored in memory; how it is recalled, used and elaborated; and how it is used in comprehension. Considerable progress has been made in the elaboration of schema theory in the past decade (see, for example, Anderson & Pearson, 1984; Brewer & Nakamura, 1984). It has played a major role in the direction of research on reading and is beginning to have an impact on the design of instructional materials. To date, however, there seems to be little, if any indication that schema theory has had an influence on the nature of standardized reading tests.

Schmata can involve both declarative and procedural knowledge and can have a strong influence on an individual's ability to comprehend and remember text. The role of schemata in facilitating comprehension was nicely illustrated by Curtis and Glaser (1983) using the following excerpt taken from Bransford and Johnson (1973).

"The procedure is actually quite simple. First you arrange

things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If your have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run this may not seem important but complications can easily arise" (Bransford & Johnson, 1973, p. 400).

The words and sentences are straight forward, but it is difficult to construct a coherent interpretation of the passage. However, if told that the passage is about washing clothes, the description fits an already available schmata and is much more readily understood.

Possession of appropriate schmata facilitates comprehension. This conclusion has useful implications for instruction, but poses a dilemma for test construction. Since individuals differ in the schemata, the prior knowledge that they bring to the test situation, the choice of passages for the test can obviously be an important determinant of the relative performance of different individuals.

Results of a study by Johnston (1981) demonstrate that differences in the match of test passages to the prior knowledge of test takers can have a substantial effect on performance. Johnston constructed a test based on two passages: one dealing with the specialization of corn in the United States and one dealing with the financial problems of the Chicago Regional Transit Authority (RTA). Both passages were administered to a sample of students in a rural community and to a sample from an urban community. As predicted, the rural sample outperformed the urban sample on the test based on the corn passage, whereas the converse was true for the RTA passage.

Although not explicitly as a consequence of schema theory, test constructors have attempted to avoid, or more accurately, to minimize the effects of prior knowledge on reading test performance. Topics assumed to require specialized prior knowledge are avoided. The primary way in which the effects of prior knowledge are minimized, however, is by the use of multiple passages which are selected in ways that attempt to balance the relative advantages and disadvantages of different groups of test takers. One of the consequences of this approach, however, is that tests typically consist of a number of short passages on different, sometimes obscure, topics that hold little motivation or interest for the reader. Furthermore, this approach does more to conceal than to eliminate the effects of prior knowledge (Valencia & Pearson, 1986).

The reading test of the Elementary level of the Metropolitan Achievement Tests, Form JS, for example, contains 11 passages ranging from 89 to 148 words in length. The passages deal with such varied topics as earthworms, bakeries, maple syrup, a turtle race, and a fanciful story about a wood chopper.

Short passages that are typical of reading tests place fewer cognitive demands for the formation of an intergrated representation of the information than do longer segments found in textbooks and other reading materials that students are expected to read and understand. This suggests that longer text segments may need to be used. Of course, feasibility would dictate that if longer passages passages were used, they would have to be fewer in number. As a consequence, such tests might be less fair than existing tests because they would lack the same degree of balance across a number of topics

that is made possible by the use of many short passages.

Nonetheless, strong arguments in favor of longer passages can be made. Longer passages can be made more comparble to the important reading tasks that students are asked to perform in school. They can involve more meaningful and coherent stories or pieces of expository text. Thus, some reading researchers (e.g., Pearson, personal communication) have argued that alternative approaches to the potential problems of bias that may be introduced by the use of a small number of longer passages should be sought.

The solution to the dilemma may depend on the creation of techniques for assessing relevant prior knowledge prior to administering a reading test. This was attempted with some success by Johnston (1981) using passage specific vocabulary test items. Several other possibilities have been suggested, some of which are currently under investigation. For example, Curtis and Glaser (1983) have suggested that comparisons of performance in reading and listening comprehension as another possible approach to this problem.

Pearson and Valencia are currently conducting research for the Illinois Assessment Program in which several approaches to the assessment to prior knowledge are being investigated. Using a variety of response formats, they ask children to make predictions about what they are likely to find in a passage on a given topic or to indicate which of a number of concepts are apt to be closely or somewhat related to the topic of the passage. While it is too early to know which techniques will prove most effective, this work is important to the development of reading tests that involve longer passages with greater ecological validity that also take into account individual differences in prior knowledge.

Whatever the nature of the passages, measurement of passage comprehension generally depends on questions that require the test taker to make inferences. Published reading tests use a variety of approaches in attempting to measure a test taker's ability to draw inferences from text. Examples include items that ask the test taker to select the best title for a story, to indicate the author's intent or point of view, or to make inferences about the feelings of a character or likely explanations of events.

The following illustrations of questions that are intended to assess the ability of children to draw inferences from a passage were taken from Level E of the CTBS, Form U.

"Which of these would most likely be a Norman Rockwell picture?"

"Why was this animal called Thunder Lizard?"

"What is this story mostly about?"

Some similar examples from the Advanced 1 Level of the Metropolitan, Form JS are:

"The best title for this story is - "

"The author feels that city play streets are -"

"At the end of this story, the actions of the lovers were motivated by their feeling of -"

It would appear that none of the above illustrative questions can be answered by simply matching surface features of the question and the passage. They all require some degree of inference to be made based on the test taker's comprehension of the passage. However, answers can sometimes be determined by simple associations and the types of inferences required are fairly limited. Moreover, because of the nature of the short test passages and because the text is avaliable

for reference in answering the questions, such questions do not exhaust the range of integrative and inferential skills required in many real-world reading tasks.

As noted in the previous section, for example, an integrated representation of a text enables a reader to identify inconsistencies in text. Markman (1979) argued for the importance of inconsistency detection as follows: "to notice inconsistencies children have to encode and store the information, draw the relevant inferences, retrieve and maintain the (inferred) propositions in working memory, and compare them" (p. 643). Though seldom used on reading tests, an approach that requires a test taker to identify inconsistencies in a text may provide a powerful means of assessing the degree to which information from different parts of text is integrated into a coherent representation of the message. Care would obviously need to be taken, however, that inconsistencies are crucial to understanding important aspects of the text.

The fact that a test taker can refer back to the text when responding to test questions, especially text-explicit ones, reduces the demand to develop an integrated representation of the passage. This suggests that it may be desirable "to prevent the reader from referring to the text while answering the questions" (Johnston, 1984, p. 236).

Another alternative approach to testing that has received increasing attention in the past few years is typified by the work of Campione and Brown (1985) on dynamic assessment. The emphases in dynamic assessment are on change that occurs during relatively brief periods of guided instruction and on the ability to transfer what is learned to new situations. A series of studies reviewed by Campione

and Brown (1985), indicate that the best predictors of future learning are not static tests, but the changes that take place during the dynamic assessment process and measures of a student's ability to "transfer their newly learned skills to novel situations" (p. 35). Although the types of tasks (e.g., Raven Pgrogressive Matrices tasks) that have been most used in research on dynamic assessment are quite different from those needed to assess reading comprehension, the approach is certainly worthy of careful consideration.

Conclusion

There is a relatively poor match between current theory and experimental research on the reading process and existing standardized tests of reading. This lack of correspondence may be due, in part, to a difference in goals and, in part, to a lag between research and practice. Survey tests have long followed a traditional psychometric perspective derived from the goals of ranking and sorting students based on individual differences in performance. Normative comparisons and predictive validity have been the hallmarks of this approach. Within this tradition, current standardized tests are relatively effective and quite efficient. The distinctions among the interrelated components of reading articulated by Curtis and Glaser (1983) which have provided the framework for this review, are of little importance to the traditional goals of norm-referenced tests.

Diagnostic tests and criterion-referenced tests have followed a different tradition. Yet the match between these tests and the theoretical perspective emphasized in this report is equally poor. Such tests generally conceive of reading as a host of hierarchically organized discrete skills. Little emphasis is given to the

integrative processes involved in reading.

It remains to be seen whether tests derived from modern cognitive theories of the reading process will lead to substantial improvements in reading assessment. Certainly, it seems unlikely that substantial improvements in prediction will result. However, prediction is not a primary goal for such tests. Providing more instructionally relevant information is. As Curtis and Glaser (1983, p. 144) concluded, "[a] combined enterprise representing test design based on knowledge of human learning and performance, psychometric requirements, and studies of test use should improve our ability to link testing and instruction."

References

Adams, M. J. & Huggins, A. W. (1985). The growth of children's sight vocabulary: A quick test with educational implications. (Technical Report No. 330). Champaign, IL: University of Illinois, Center for the Study of Reading.

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42, 145-170.

Anderson, R. C. (1977). The notion of schemata and the educational enterprise: General discussion of the conference. In R. C. Anderson, R. J. Spiro, & W. E. Montague (eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum.

Anderson, R. C. & Freebody, P. (1981). Vocabulary knowledge (Technical Report No. 136). Champaign, IL: University of Illinois, Center for the Study of Reading.

Anderson, R. C. & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. Advances in Reading/Language Research. 2, 231-256.

Anderson, R. C. & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson (Ed.), Handbook of reading research. New York: Longman.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Framework for comprehending discourse. American Educational Research Journal, 14, 367-381.

Biemiller, A. (1977-78). Relationships between oral reading rates for letters, words and simple test in the development of reading achievement. Reading Research Quarterly, 13, 223-253.

Biemiller, A (1981). Biemiller Test of Reading Processes. Toronto,

Ontario: University of Toronto, Guidance Center.

Bransford, J. D. & Johnson, M. K. (1973). Consideration of some problems in comprehension. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press.

Brewer, W. F. & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & T. K. Srull (Eds.), Handbook of social cognition (Vol. 1). Hillsdale, NJ: Erlbaum.

Brown, A. (1978). Knowing when, where and how to remember: A problem of matacognition. In R. Glaser (Ed.), Advances in instructional psychology, Vol. 1. Hillsdale, NJ: Erlbaum.

Campione, J. C. & Brown, A. L. (1985). Dynamic assessment: One approach and some initial data. (Technical Report No. 361). Champaign, IL: University of Illinois, Center for the Study of Reading.

Carroll, D. J. (1985). Review of the Instant Words Criterion Test. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Carroll, J. B., Davies, P., & Richman. B. (1981). Word frequency book. New York: American Heritage Publishing Co.

CTB/McGraw-Hill. (1982). Comprehensive Tests of Basic Skills, Forms U and V. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1980). PRI Reading Systems. Monterey, CA: CTB/McGraw-Hill.

Curtis, M. E. (1980). Development of components of reading skill. Journal of Educational Psychology, 72, 656-669.

Curtis, M. E. & Glaser, R. (1983). Reading theory and the assessment of reading skill. Journal of Educational Measurement, 20, 133-147.

Dale, E. & O'Rourke, J. (1976). The living word vocabulary - The words we know. Elgin, IL: Dome Press, Inc.

Draper, A. G. & Moeller, G. H. (1971). We think with words (therefore, to improve thinking, teach vocabulary). Phi Delta Kappan, 52, 482-484.

Ekwall, E. E. (1979). Ekwall Reading inventory. Boston: Allyn & Bacon.

Elkins, J. (1985). Review of the Instant Words Criterion Test. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Flavell, J. J. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

Freebody, P. & Anderson R. C. (1981). Effects of differing proportions and locations of difficult vocabulary in text comprehension. (Technical Report 202). Champaign, IL: University of Illinois, Center for the Study of Reading.

Fry, E. (1980). The Instant Words Criterion Test. Providence, RI: Jamestown Publishers.

Fry, M. A. (1985). Review of Biemiller Test of Reading Processes. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Gates, A. I., McKillop, A. S., & Horowitz, E. C. (1981). Gates-McKillop-Horowitz Reading Diagnostic Test, Second edition. Hagerstown, MD: Teachers College Press.

Guttentag, R. E. & Haith, M. M. (1978). Automatic processing as a function of age and reading ability. Child Development, 49, 707-716.

Hanna, G., Schell, L. M., & Schreiner, R. (1977). The Nelson Reading Skills Test, Forms 3 and 4. Chicago, IL: Riverside Publishing Co.

Hiebert, E. H. (1985). Review of PRI Reading Systems. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln,

NE: The Buros Institute of Mental Measurements.

Hogaboam, T. & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. Journal of Educational Psychology, 70, 717-729.

IOX Assessment Associates. (1980). The IOX Basic Skills Word List. Los Angeles, CA: IOX Assessment Associates.

Jastic, J. F. & Jastic, S. (1978). Wide Range Achievement Test, 1978 edtition. Willimington, DE: Jastic Associates.

Jenkins, J. R., Pany, D., & Schreck, J. (1978). Vocabulary and reading comprehension: Instructional effects. (Technical Report 100). Champaign, IL: University of Illinois, Center for the Study of Reading.

Johns, J. E. (1981). Basic Reading Inventory, Second edition. Dubuque, IA: Kendall/Hunt.

Johnston, P. H. (1981). Prior knowledge and reading comprehension test bias. Ph.D. dissertation. University of Illinois at Urbana-Champaign.

Johnston, P. H. (1983). Reading comprehension: A cognitive basis. Newark, DE: International Reading Association.

Johnston, P. H. (1984). Prior knowledge and reading comprehension test bias, Reading Research Quarterly, 19, 219-239.

Juel, D. (1980). Comparison of word identification strategies with varying context, word type, and reader skill. Reading Research Quarterly, 16, 358-376.

Juel, D. & Roper-Schneider, D. (1985). The infuluence of basal readers on first grade reading. Reading Research Quarterly, 20, 134-152.

Koslin, B. L., Koslin, S. & Zeno, S. (1979). Towards an effectiveness measure in reading. In R. W. Tyler & S. H. White (Eds.), Testing,

teaching, and learning: Report of a conference on research on testing. Washington, DC: National Institute of Education.

Lesgold, A. M. & Curtis, M. E. (1981). Learning to read words efficiently. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, NJ: Erlbaum.

Lesgold, A. M. & Perfetti, C. A. (1977). Interactive process in reading. Discourse Process, 1, 323-336.

MacGintie, W. H. (1978). Gates-MacGintie Reading Tests. Chicago, IL: Riverside Publishing Co.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. Child development, 48, 986-992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

McConkie, G. W. & Zola, D. (1991). Language constraints and the functional stimulus in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), Integrative processes in reading. Hillsdale, NJ: Erlbaum.

McCormick, C. & Samuels, S. J. (1979). Word recognition by second graders: The unit of perception and interrelationships among accuracy, latency and comprehension. Journal of Reading Behavior, 11, 107-118.

National Academy of Education. (1985). Becoming a nation of readers: A report of the Commission on Reading. Washington, DC: National Institute of Education.

Pearson, P. D. (1986). Twenty years of research on reading comprehension. In T. E. Raphael (Ed.), The context of school based literacy. New York: Random House.

Pearson, P. D. & Herman, P. (1985). Review of Gates-McKillop-Horowitz

Reading Diagnostic Test, Second edition. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Pearson, P. D. & Spiro, R. J. (1980). Toward a theory of reading comprehension instruction. Topics in Language Disorders, 1, 71-88.

Perfetti, C. A., Finger, E., & Hogaboam, T. (1978). Sources of vocalization latency differences between skilled and less skilled young readers. Journal of Educational Psychology, 70, 730-739.

Perfetti, C. A. & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Perfetti, C. A. & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), Cognitive process in comprehension. Hillsdale, NJ: Erlbaum.

Perfetti, C. A. & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol 1.). Hillsdale, NJ: Erlbaum.

Plessas, G. P. (1985). Review of Basic Reading Inventory, Second edition. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Linclon, NE: The Buros Institute of Mental Measurements.

Pope, L. (1974). Pope Inventory of Basic Reading Skills. North Bergin, NJ: Book Lab, Inc.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1978). Meteropolitan Achievement Tests, Fifth edition. San Antonio, TX: The Psychological Corporation.

Rumelhart, D. E. & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Sprio, & W. E. Montague (Eds.),

Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum.

Rupley, W. H. (1985). Review of Gates-MacGinitie Reading Tests. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Sakiey, E. & Fry, E. (1979). 3,000 instant words. Providence, RI: Jamestown Publishers, Inc.

Samuels, S. J. & LaBerge, D. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Shankweiler, D. & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanaugh & I. G. Mattingly (Eds.), Language by ear and eye. Cambridge, MA: The MIT Press.

Smith, C. B. & Arnold, V. A. (1983). Testing and management resource book . New York: Macmillan.

Spache, G. D. (1981). Diagnostic Reading Scales. Monterey, CA: CTB/McGraw-Hill.

Stanovich, K. E. (1980). Toward an integrative-compensatory model of individual differences in the development of reading fluency. Reading Research Quarterly, 16, 32-71.

Taylor, S. E., Frackenpohl, H., White, C. E., Nieroroda, B. W., Browning, C. L., & Birsner, P. E. (1979). EDL core vocabularies in reading, mathematics, and social studies. New York: McGraw-Hill.

Thorndike, R. L. (1973). Reading comprehension education in fifteen countries. New York: Wiley.

Touchstone Applied Science Associates (1983). Degress of Reading Power. New York: The College Board.

Tuinman, J. J. (1974). Determining the passage dependencey of comprehension questions in 5 major tests. Reading Research Quarterly,

9, 206-223.

Tuinman, J. J. & Brady, M. E. (1974). How does vocabulary account for variance in comprehension test? A preliminary instructional analysis. In P. Nacke (Ed.), Twenty-third yearbook of the National Reading Conference. Clemson, SC: National Reading Conference.

Valencia, S. & Pearson, P. D. (1986). Reading assessment: Time for change. (unpublished manuscript). Champaign, IL: University of Illinois, Center for the Study of Reading.

Weinstein, M. A. (1983). Testing and management resource book: Blackline masters, Grade 3. New York: Macmillan Publishing Co.

West, R. F. & Stanovich, K. E. (1979). The development of automatic word recognition skills. Journal of Reading Behavior, 11, 211-219.

West, R. F. & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. Child Development, 49, 717-727.